



INTEGRATIVE TESTING OF HOW ENVIRONMENTS FROM THE PAST TO THE PRESENT SHAPE GENETIC STRUCTURE ACROSS LANDSCAPES

Qixin He,¹ Danielle L. Edwards,² and L. Lacey Knowles^{1,3}

¹Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

²Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520

³E-mail: knowlesl@umich.edu

Received September 7, 2012

Accepted April 26, 2013

Data Archived: Dryad doi:10.5061/dryad.77077

Tests of the genetic structure of empirical populations typically focus on the correlative relationships between population connectivity and geographic and/or environmental factors in landscape genetics. However, such tests may overlook or misidentify the impact of candidate factors on genetic structure, especially when connectivity patterns differ between past and present populations because of shifting environmental conditions over time. Here we account for the underlying demographic component of population connectivity associated with a temporarily dynamic landscape in tests of the factors structuring population genetic variation in an Australian lizard, *Lerista lineopunctulata*, from 24 nuclear loci. Correlative tests did not support significant effect from factors associated with a static contemporary landscape. However, spatially explicit demographic modeling of genetic differentiation shows that changes in environmental conditions (as estimated from paleoclimatic data) and corresponding distributional shifts from the past to present landscape significantly structures genetic variation. Results from model-based inference (i.e., from an integrative modeling approach that generates spatially explicit expectations that are tested with approximate Bayesian computation) contrasts with those from correlative analyses, highlighting the importance of expanding the landscape genetic perspective to tests the links between pattern and process, revealing *how* factors shape patterns of genetic variation within species.

KEY WORDS: Coalescent, demographic simulation, gene flow.

Although temporal scale is one of the primary distinguishing factors of landscape genetic and phylogeographic study, such a distinction is not only unnecessary, but also potentially problematic. For example, landscape genetics studies how contemporary habitat suitability and connectivity influence population genetic structures spatially (Manel et al. 2003; Storfer et al. 2007). Phylogeography typically focuses on historical processes that generated the patterns of genetic variation (Avise et al. 1987; Knowles 2009). There may certainly be cases in which one of the two processes predominates (e.g., Knowles et al. 2007; Hull et al. 2008; Xu et al. 2009; Mendez et al. 2010; Perrier et al. 2011). Yet, because such

studies are often pursued under one of the two perspectives, their joint influence can be overlooked, risking the misidentification of factors structuring patterns of genetic variation.

As both landscape genetic and phylogeography shift toward the analysis of multilocus data, and specifically as next-generation sequencing technologies become widely applied (e.g., Gompert et al. 2010; Thomson et al. 2010), concerns over molecular markers as a distinguishing factor between landscape genetics and phylogeography (e.g., Wang 2010) will certainly diminish. Similarly, the greater power and resolution provided by such data sets opens up new possibilities for expanding methodologies that can

test causation of, as opposed to seeking associations with, the underlying patterns of genetic variation.

The melding of disciplines is represented in the approach advocated here, which we illustrate with an empirical example—specifically, a test aimed at revealing *how* geography and the environment shape patterns of genetic variation in a lizard, *Lerista lineopunctulata*. This lizard is distributed along the southwestern Australian coastal sand plains or dunes (Fig. 1; Cogger 2000; Wilson and Swan 2008). Sea-level changes in glacial and interglacial periods expanded or contracted suitable coastal sand habitats for the species (Storr and Harold 1978; Hocking et al. 1987). Consequently, it is conceivable that population divergence could reflect the contemporary habitat configuration, which limits migration among the small geographically isolated populations (Excoffier et al. 2009), or colonization associated with historical shifts in the species distribution (Zellmer and Knowles 2009), given that a habitat specialist would track climate-induced habitat shifts. We first conduct both individual- and population-level correlative tests to identify potential factors structuring genetic variation, including geography, climatic, and soil characteristics (see also Edwards et al. 2012). We then move beyond these traditional descriptive landscape genetic analyses (Legendre and Fortin 2010) with an approach that provides quantitative species-specific predictions that account for the interaction between abiotic and biotic factors (i.e., the environmental factors and the life-history characteristics of taxa that mediate the impact of these factors on survival and movement patterns; see Knowles and Alvarado-Serrano 2010; Brown and Knowles 2012). Specifically, we generated a large multilocus data set to test whether the current genetic structure reflects (i) the geographic configuration of populations; (ii) the contemporary environment; or (iii) the dynamic history of shifting environmental characteristics since the last glacial maximum.

Our work highlights the potential synergy between traditional landscape genetic approaches and model-based inferences by translating hypotheses identified from correlative analyses into a suite of alternative demographic processes that can be formulated as models (see also Bruggeman et al. 2010; Epperson et al. 2010; Landguth et al. 2010; Morgan et al. 2011; Shirk et al. 2012). Our approach contrasts with the tradition of intuiting qualitative phylogeographic hypotheses from ecological niche models (ENMs; reviewed in Knowles 2009). Here quantitative information about variation in the habitat suitabilities across space and time is used to inform a spatially explicit demographic model whose parameters are then used for coalescent simulations. As a consequence, predicted patterns of genetic variation are species specific, reflecting the interaction between the physical environment and biological parameters (e.g., local population sizes and migration rates) that determines the level and pattern of gene flow across the landscape (see Knowles

and Alvarado-Serrano 2009; Morgan et al. 2011; Brown and Knowles 2012). In addition, we rigorously test these models using approximate Bayesian computation (ABC; Beaumont et al. 2002), and assess the quality of parameter estimates using pseudo-observed datasets, *pods* (see Bertorelle et al. 2010; Robert et al. 2011).

With reference to the empirical study of *L. lineopunctulata*, we highlight how extrapolating causation from descriptive correlates of genetic variation with the environment and geography would be misleading (see also Meirmans 2012), but was avoided by applying model-based inference with an expanded repertoire of models (i.e., not only isolation-by-distance [IBD], but also models that include additional environmental factors, and temporal shifts in habitats across the landscape). This approach, iDDC modeling, integrates distributional, demographic, and coalescent models to generate predictions for species-specific patterns of genetic variation. With the intent that the methods proposed here can be generally applied to different biological systems that had experienced nonstatic demographic history, we include a discussion of not just the promise of iDDC modeling (see approaches described in Ray et al. 2005; Neuenschwander et al. 2008), but also the limitations.

Methods

SAMPLING AND MOLECULAR DATA

Lerista lineopunctulata tissue samples ($N = 89$) were field collected or obtained from the Western Australian Museum and Australian Biological Tissue Collections (South Australian Museum; Table S1) for full geographic coverage of the species, with multiple individuals sampled from each of the delimited populations (see Edwards et al. 2012, for details about population assignment). Note that the southern populations, formerly assigned to *L. lineopunctulata*, are considered a separate species under taxonomic revision (D. L. Edwards, P. Doughty, and J. S. Keogh, unpubl. data) and have not been included in this study (see also Edwards et al. 2012). Also note that the number of loci in this study was expanded considerably from 3 to 24 (a prerequisite for testing hypotheses, as opposed to describing genetic variation, as in Edwards et al. 2012).

Anonymous nuclear loci were developed from a Roche 454 sequencing run (procedures similar to Bertozzi et al. 2012), and supplemented with sequences for loci from published primers (see Table S2). Marker development from a 454 run used one individual of the focal taxon, *L. lineopunctulata*, and one individual of *Lerista praepedita*. Note that *L. praepedita* was used to identify variable markers while avoiding ascertainment bias that results from using intraspecific screening sets (see Carstens and Knowles 2006). Details regarding preparation of DNA samples for developing markers are given in Gompert et al. (2010). This

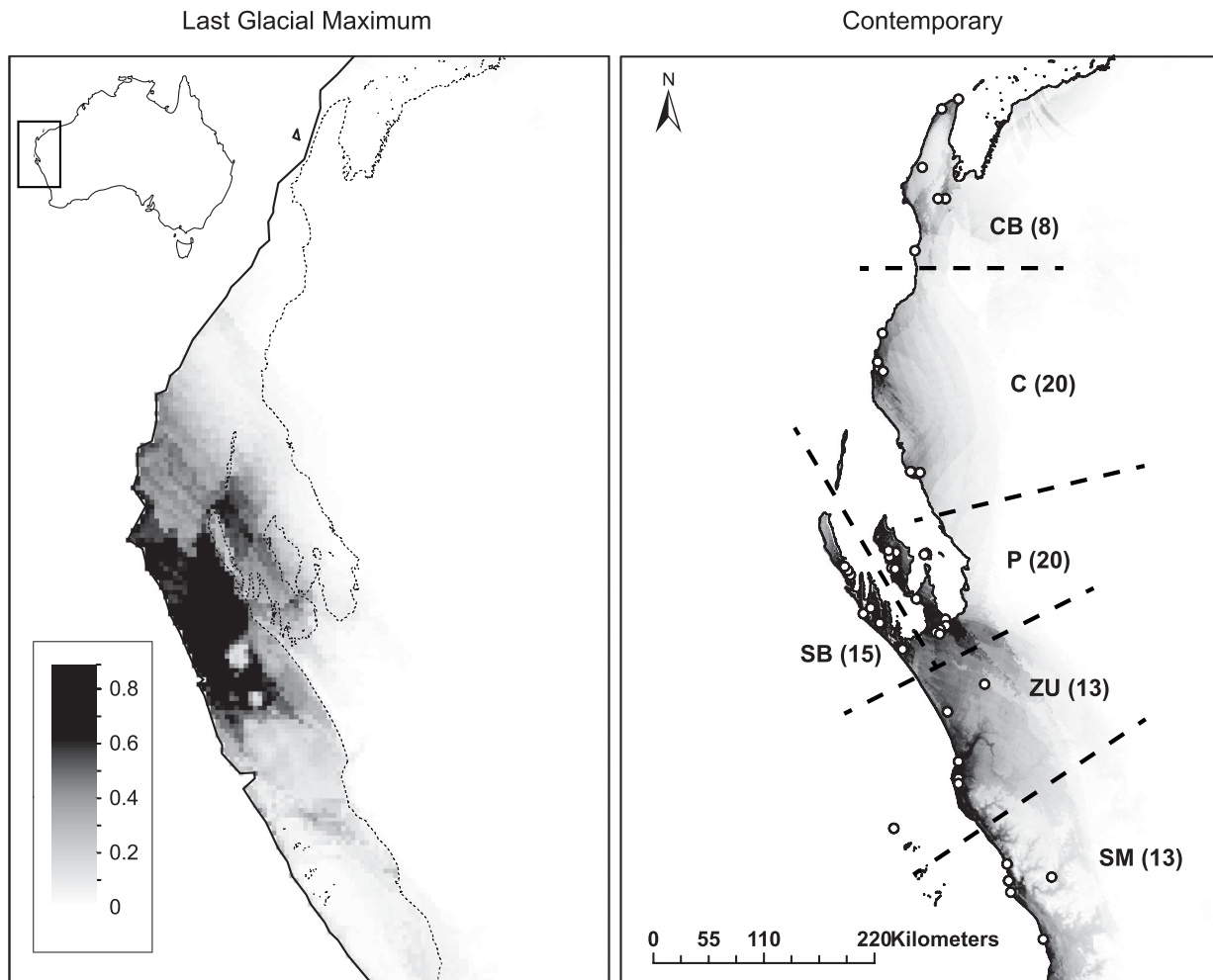


Figure 1. Predicted contemporary and past distribution of *Lerista lineopunctulata* in southwest Australia (see inset for location in continent) based on climatic and paleoclimatic variables, respectively (see text for details). Habitat suitability scores are shown as ranging from the lowest (lightest) to the highest (darkest) suitability. Dashed lines separate populations (as determined from barriers associated with breaks in suitable habitat; see Edwards et al. 2012) and population names along with sample sizes (in parentheses) are shown with dots that mark sampling sites. In contrast to the linearly distribution of suitable habitat along the coast today, refugial areas for the species 21 kya were more circumscribed and extended westward of current populations SB and P (dashed outline marks the current coast line), given the emergence of vast areas of coastal sand habitats during glacial maximum (Hocking et al. 1987; Mory et al. 2003).

entailed the construction of a reduced representation library for each species from genomic DNA digested with *EcoRI* and *MseI* enzymes. Unique barcodes were ligated for each species and size-selected fragments in equimolar concentrations were used for the Roche 454 sequencing. The sequences were trimmed and quality filtered using custom perl scripts and assembled using the NGen sequence assembler version 2.0 (DNASTAR); settings used in the assembly are provided in Table S3. Contig consensus sequences were screened against BLAST to ensure loci were not mtDNA or transposable elements and did not belong to known gene families. Primers were designed for amplifying and sequencing fragments between 150 and 700 bp using traditional Sanger-sequencing at the University of Michigan DNA core facility.

A total of 24 nuclear loci were sequenced in both directions in each individual (GenBank KC545970–KC549439), although there were some missing data due to polymerase chain reaction (PCR) failures (Table S4). Eighteen loci were identified from the 454 run that produced clear bands, with single-copy sequences and contained at least one variable site between the two species (*L. lineopunctulata* and *L. praepedita*). In addition, we sequenced six loci using published primers (see Table S2 for references). PCR reactions were run in 20 μ L volumes with 2 μ L 10 \times reaction buffer, 0.8–2.5 μ L 50 mM $MgCl_2$, 1 μ L 10 mM dNTPs, 0.4 μ L bovine serum albumin (10 mg/mL), 0.8 μ L of each 10 μ M primer, 1U Taq polymerase, \sim 100 ng gDNA and the volume made up to 20 μ L with ultra-pure H_2O .

PCR cycles were 95°C 1 min; 30 cycles of 95°C 30 sec, 59–65°C 20 sec, 72°C 45 sec; 72°C 4 min (see Table S2 for specific conditions and exceptions). Haplotype phase was determined using PHASE (Scheet and Stephens 2006).

SPECIES ECOLOGICAL NICHE MODELING

In addition to our collected samples, occurrence data of *L. lineopunctulata* were collated from OZCAM (www.ozcam.org.au) and geo-referenced (see Edwards et al. 2012). Projections of current species distributions based on habitat suitability was estimated from 19 current climate layers from the WorldClim global climate database (www.worldclim.org). The ecological niche models (ENMs) were generated with MaxEnt version 3.3.3k (Phillips et al. 2006) from 10 cross-validation runs, which accurately predicted the species distribution (area under the curve [AUC] value of 0.971). The model was also used to predict the past distribution of the species at the Last glacier maximum (LGM) using the same 19 climate layers from the Community Climate System Model derived from PMIP2 database available on WorldClim database (Hijmans et al. 2005).

TESTS OF ASSOCIATIONS WITH GENETIC STRUCTURE

The potential impact of environmental factors on genetic structures was tested at both the individual and population level. Distance-based redundancy analysis (dbRDA; Legendre and Anderson 1999) was used to test for the relationship between individual pairwise genetic distances and corresponding climatic and soil variables (i.e., the score at the sampling site where the individual was collected), conditioned on geographic distances (i.e., removing the effect of geographic distance separating individuals). Distance-based redundancy analysis is a multivariate technique for testing a distance-based matrix (in this study, the matrix of pairwise genetic distances) against rectangular predicting variables, in which the relationship between the principal coordinates of the distance matrix and the variables are then analyzed. For analyses of an association between environmental factors and population-level genetic structure, pairwise F_{ST} -values were calculated among populations using Arlequin 3.5 (Excoffier and Lischer 2010) and the environmental differences among populations were summarized for an isolation-by-resistance test.

Individual pairwise genetic distances were calculated in Arlequin 3.5 with Tajima and Nei's correction (Tajima and Nei 1984). For this analysis, the multilocus data were condensed into two haplotypes per individual by concatenating one of the two alleles (selected randomly) for each locus across loci. Positions with more than 60% missing data (across individuals) were not included in the calculation of individual pairwise genetic distances. Calculations of environmental distances were conducted on the principal component 1 (PC1) from a principal component analysis (PCA) of the 19 climate layers extracted from ArcGIS10 due

to correlation across climatic layers (Manel et al. 2001; Hirzel et al. 2002; Peterson et al. 2011). We performed PCA directly on the climate layers instead of values extracted from sampling points because we want to capture the variation in the environment but avoid any bias in the sampling points. This climate PC1 explained 90% of the variation in the climatic data. We also characterized the spatial variation in soil characteristics. Soil properties were derived from the soil-type data in the Atlas of Australian Soils (Northcote et al. 1960) from the Australia Soil Information System (<http://www.asris.csiro.au>) and interpreted following McKenzie et al. (2000) as 13 measurements of the soil profiles including percentage of clay, thickness, water flow, nutrients (see Table S5), which were summarized with a PCA. This soil PC1 explained 85% of the total variation in the soil data and was retained for analysis with dbRDA. Pairwise Euclidean geographic distances among all individuals were calculated in ArcGIS 10. Because dbRDA only relates a matrix to rectangular predictors, the geographic Euclidean distance matrix was transformed into continuous rectangular vectors via principal coordinates analyses using the *pcnm* function of the *Vegan* package (Oksanen et al. 2012) in R (R Core Team 2012). Each of the three potential predictors (geographic distance, climate-PC1, and soil-PC1) were tested separately against genetic distance using the *capscale* function in the *Vegan* package, as well as tests of climate-PC1 and soil-PC1 conditioned on geographic distance (i.e., partitioning out the effect of geographic distances).

For the analyses of isolation-by-resistance (McRae 2006) used in the population-level tests of association between environmental factors and genetic variation, the average resistance among populations was estimated in Circuitscape version 3.5.8 (Shah and McRae 2008) using habitat suitability score as per-cell conductance. Specifically, for each population, a convex hull (i.e., a polygon) that encompassed the minimum population area from sampling localities was used to define the region from which Circuitscape calculated resistance scores to represent the connectivity among populations. Isolation-by-resistance was tested using Mantel and partial Mantel tests in IBDWS version 3.23 (Jensen et al. 2005) for population-level associations of genetic variation and environmental factors by considering the habitat suitabilities modeled from contemporary climatic variables, as well as the average habitat suitability of the current and past climatic conditions (i.e., an intermediate landscape shown in Fig. 2). We chose to use partial Mantel tests because all the explanatory factors are distance-based matrices (Legendre and Fortin 2010) and the primary interest here is on the change in correlations between the predictor matrix and the genetic distance, and therefore the issues surrounding the interpretation of *P*-values with partial Mantel tests (Raufaste and Rousset 2001) and reduced power in detecting relationships compared to dbRDA (Legendre and Fortin 2010) is not a critical problem as applied here.

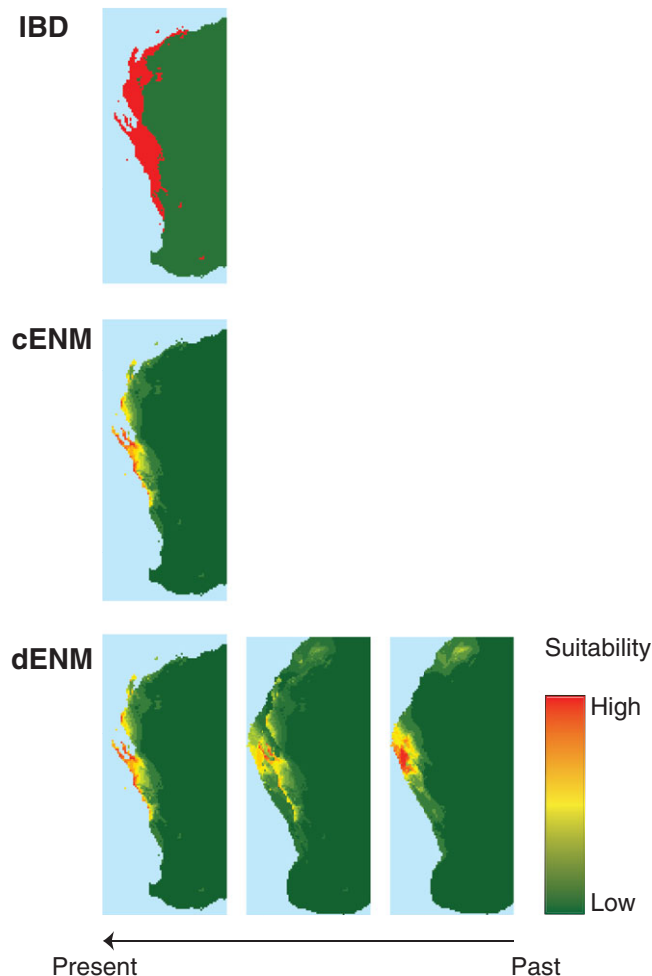


Figure 2. Schematic of the three spatially explicit models used in the demographic simulations to evaluate how environmental factors, as well as changing environmental conditions associated with the Pleistocene glaciation, might be causality related to patterns of genetic variation. For each model, variation in the underlying environmental components used for the demographic simulations is shown (see Knowles and Alvarado-Serrano 2010). The respective models are: (i) isolation-by-distance (IBD); (ii) contemporary ecological niche model (cENM); and (iii) dynamic ENM (dENM; as described in detail in the text); shown for each model is the spatially explicit layer that formed the basis for the demographic simulations. Note that both the IBD and cENM models are static in the sense that the habitat suitability scores used for the demographic modeling were the same across generations, whereas with the dENM model is dynamic with habitat suitability scores changing over time from the last glacial maximum to the present in a step-wise fashion, as shown (see Supporting Information for details). After each forward-time demographic simulation, coalescent simulations were run for sampled individuals backward in time.

INCORPORATING SPATIALLY EXPLICIT DEMOGRAPHIC HISTORY INTO MODEL TESTS WITH ABC ANALYSES

We apply the iDDC-modeling approach so that we can examine if correlations between environmental factors or historical shifts in distributions might (or might not) reflect causal relationships with the processes governing population genetic structure. Specifically, with iDDC-modeling a population demographic model is used to make explicit predictions for patterns of genetic variation (Currat and Excoffier 2004; Wegmann et al. 2006; Sork et al. 2010), where the population demography is informed by the underlying environment (i.e., it takes into account spatial and temporal heterogeneity of the environment in a species-specific manner; see details in Knowles and Alvarado-Serrano 2010; Brown and Knowles 2012). To test whether the current genetic structure results from (a) the geographic configuration of populations, (b) the contemporary environment, and (c) the dynamic history of shifting environmental characteristics associated with the differences between the present and the last glacial maximum, we constructed three corresponding demographic models and used coalescent simulations to predict genetic variations. In contrast with the studies to date utilizing iDDC modeling, we then use these simulations for identifying the most probable model and estimating parameters using approximate Bayesian computation (ABC; see Beaumont et al. 2002 for an overview of ABC).

The general procedure involves translating the habitat suitability scores from an ENM into spatially explicit population parameters for demographic simulations, which are then used for a spatially explicit coalescent simulation to generate expected patterns of genetic variation (for details about the procedures, see Knowles and Alvarado-Serrano 2010; Brown and Knowles 2012). This flow of information provides direct links between process and pattern. Specifically for this study, we statistically downscaled the maps from the ENMs for the current and past climatic conditions to 0.1 decimal degree ($\sim 121 \text{ km}^2$ per cell) to have a tractable number of demes for demographic simulation. All spatially explicit demographic simulations were performed in SPLATCHE2 (Currat et al. 2004), with population carrying capacities scaled proportionally to the local habitat suitability score (i.e., the relative values per grid cell differed depending on the predicted habitat suitability derived from the ENM generated with MaxEnt). Patterns of genetic diversity were then generated from coalescent simulations based on the specific demographic simulation (i.e., genetic variation differed across the landscape depending on the probability of coalescence and migration across demes; Excoffier et al. 2000; Currat et al. 2004). We ran 24 coalescent simulations for each demographic history corresponding to each of the 24 separate loci in the empirical data set (see Supplementary Tables), such that these independent realizations of the coalescent process generated genealogies for simulating sequence data for each

locus, where the sampled individuals from the simulated data sets matched those in the empirical data. DNA sequence data were also simulated according to the empirical DNA sampling conditions (e.g., the same gene length and amounts of missing data). Relative mutation rates among loci matched those from empirical π estimations (Table S6).

The three models tested here were selected to test hypotheses motivated by the correlative analyses described earlier (Fig. 2). Specifically, the hypotheses tested were that patterns of genetic variation reflect: (i) genetic drift associated with the geographic configuration of habitats—tested using a model of IBD; (ii) genetic drift associated not only with the geographic configuration of habitats, but also differences in local population sizes and the amounts of gene flow as defined by the suitabilities of contemporary environment—tested using a model of the contemporary ENM (cENM); and (iii) genetic drift associated with distributional shifts caused by changes in environmental conditions—tested using a dynamic ENM (dENM) model. These models again differ with respect to input layers used for the demographic simulations (see Fig. 2). Note that the cENM model considers the impact of habitat heterogeneity on patterns of genetic variation, whereas the IBD model only considers the influence of geographic distance, but both of these models are “static” models in that the layer informing the demographic model does not change over time. In contrast, the dENM also considers how a shifting distribution, and the accompanying colonization process, impacts patterns of genetic variation (Fig. 2). Temporal variation in habitat suitability was modeled in a step-wise fashion (i.e., using the habitat suitability scores from three period-specific ENMs; see also Brown and Knowles 2012). Specifics regarding the simulation details for all the models are given in the Supplemental Methods. Note that each generation during the demographic simulation, m proportion of the population migrates out of the local deme; migration occurs in the adjacent four cells (north, south, west, east) and the allocation to different directions are defined by the friction score (see Supplemental Methods); after exchange of individuals, populations grow logistically at the rate of 1 regulated by the carrying capacity inferred from habitat suitability. Ecological niche model maps and the settings for demographic modeling in Splat2 are deposited in Dryad.

Model selection and parameter estimation were conducted using ABC with ABCestimator in ABCtoolbox (Wegmann et al. 2010). We performed 1,000,000 simulations for each model under a standard ABC rejection sampling approach (Tavare et al. 1997; Beaumont et al. 2002). In addition to comparisons of the performances of different models, we also estimated four critical demographic/mutation parameters: maximum carrying capacity (K_{\max}), migration rate (m), ancestral population size prior to expansion (N_{Anc}), and average mutation rate (μ) because each model would have different estimates of these parameters that generate

simulated data closest to empirical ones. The ABC inference was based on a total of 34 summary statistics calculated within, between, and across all populations using Arlequin (see Table S7 for the full list of summary statistics). They include segregating sites S for each population and across populations, private segregating sites for each population PrS , the mean number of pairwise genetic differences of each population π , and pairwise population F_{ST} (Weir and Cockerham 1984). To remove the effects of interactions between summary statistics, as well as reduce “the curse of dimensionality” (i.e., when too many statistics are included, the distance between the simulated and empirical values systematically increases, reducing the accuracy of parameter estimates and making it more difficult to distinguish among models), partial least squares (PLSs) components (Boulesteix and Strimmer 2007) were extracted from all predictor variables. This treatment extracts orthogonal components from data with high dimensionalities while maximizing the covariance of summary statistics and the parameters of interests (Wegmann et al. 2009; Wegmann and Excoffier 2010). Partial least squares were calculated in the “PLS” package (Mevik and Wehrens 2007) with boxcox treatment (Box and Cox 1964) in R for the first 10,000 runs for each model. The root mean squared error (RMSE) prediction of each parameter was examined before deciding upon the number of PLS components to be used (see Fig. S1).

Five thousand simulations (0.5%) that were closest to the empirical observation were retained from each model for model selection. Postsampling regression adjustment were applied using ABC-GLM (general linear model) function (Leuenberger and Wegmann 2010) to obtain posterior distributions of the parameters, which assumes that the accepted PLS is produced by a GLM from the parameters. We use Bayes factors for model selection, which is the ratio between marginal densities of two models. The higher the ratio is, the more supported the first model is. Under the GLM model, the likelihood of the empirical data (i.e., the observation) can be evaluated and compared with the likelihoods of other retained simulations. The fraction of simulations that have a smaller likelihood than the empirical data was shown as P -value to check if the model is capable of generating the observed data. Very small P -values indicate that a model is highly unlikely (Wegmann et al. 2010). Coefficient of variation (R^2) of each parameter explained by the six used PLSs was computed as an indicator for the power of estimation (Neuenschwander et al. 2008). After selecting the highest supported model, we validated the accuracy of parameter estimation in the most supported model. A total of 1000 pseudo-observations were generated from prior distributions of the parameters. If the estimation of the parameters is unbiased, posterior quantiles of the parameters from pseudo runs should be uniformly distributed in [0,1] (Cook et al. 2006; Wegmann et al. 2010). The posterior quantiles of true parameters for each pseudo run were also calculated based on the

posterior distribution of the regression adjusted 5000 simulations closest to the pseudo-observation. Average RMSE of the model estimates for parameters of pseudo-observations was calculated to check for the accuracy of estimation.

Results

The anonymous nuclear markers were all variable (see Table S6 for summaries of molecular variation), but differed in the mutation rates (θ_{π} ranges from 0.009 to 0.1). These per-locus differences were incorporated into all correlative tests and simulations used to test hypotheses about the link between patterns and process with ABC.

ASSOCIATIONS BETWEEN PATTERNS OF GENETIC DIVERGENCE AND ENVIRONMENTAL FACTORS

A significant association between geographic distance and genetic differentiation was detected with both individual-level and population-level analyses (i.e., results from dbRDA and F_{ST} analyses, respectively). Specifically, tests of IBD with dbRDA explained 59% of the genetic variation among individuals (Table 1). A strong geographic signal was also evident from the regression of linearized F_{ST} -values against pairwise Euclidean distances between populations (Fig. 3A).

Contemporary climatic differences are also significantly associated with patterns of genetic divergence; however, when conditioned on the geographic distances between individuals (i.e., controlling for the effects of geographic isolation), the effects are not significant (Table 1). For example, although PC1 of the climatic variables was significant when tested alone, when conditioned on the geographic distance separating individuals, it was not, and the proportion of genetic variation explained decreased from 9% to 2% (Table 1). Soil-PC1 was not significant irrespective of conditioning on geography. For tests of associations between population-level divergences and environmental differences separating populations as measured by an analysis of isolation-by-resistance (McRae 2006) from habitat suitability scores for per-cell conductance, a significant association is detected (Fig. 3B, C). However, when controlling for geography with a partial Mantel tests (Table 2), the genetic differentiation actually shows an inverse relationship, as measured by the resistance from the current ENM alone, where genetic differentiation was greater for lower resistance (rather than a positive relationship between genetic differentiation and levels of resistance). We discuss this enigmatic pattern later, but further analyses suggest it could reflect the confounding influence of past environmental conditions. For example, when pairwise F_{ST} -values are regressed against pairwise resistance-scores calculated from a composite ENM map (i.e., the average habitat suitability scores from the

ENMs of past and current climatic conditions; see Fig. 2), correlation coefficients are much higher than considering only the current climate ($r = 0.84$ vs. $r = 0.46$; see Table 2), and remain significant if controlling for the effect from current climatic conditions (Table 2) using a partial mantel test. However, the highest correlation is with the average pairwise Euclidean geographic distances separating populations (Table 2), and the effect of the averaged habitat suitability over time (i.e., from past and current ENM) is not significant after controlling for the influence of Euclidean geographic distances. Yet, it would seem highly unlikely that the Euclidean geographic distance among individuals or populations reflect dispersal patterns given the shape of the coastline and the distribution of species (i.e., the animals would have to traverse inhospitable habitat, including the ocean, especially for populations P and SB; Fig. 1). This raises the question of whether the association of geography and genetic divergence actually arose under an IBD model? Second, would a model of the population demography produce patterns of genetic variation that are likely to have arisen under IBD? Finally, even though the isolation-by-resistance takes into account possible paths (McRae 2006), it does not take into account the demographic consequences of moving through the habitat. Consequently, could environmental factors (either present or past) actually impact patterns of genetic variation, but go undetected with correlative tests? The answers to these questions, which again are motivated by the aforementioned correlative analyses, are discussed in the following section.

TESTS OF THE LINKS BETWEEN PATTERN AND PROCESS

For the ABC analyses, we selected the first six PLSs for calculating the distance between simulations and the empirical observation because RMSE of the four parameters in four models does not decrease significantly with additional PLSs (see Fig. S1). Based on the marginal density for each model calculated from the 5000 closest simulations for each model, the dENM model—the model of the colonization history under dynamic ENMs (Fig. 2)—best explains the patterns of genetic divergence observed within *L. lineopunctulata*. The two static models that only consider aspects of the current landscape, the IBD and cENM, have significantly lower marginal densities. For example, even though the cENM has much higher support than the IBD model, the difference in Bayes factors between the cENM and dENM is more than 300 (a substantial difference; Jeffreys 1961). Moreover, the dENM model has a high P -value, suggesting a significant correspondence between the observed empirical data and the simulated data under this model, whereas the P -values for the cENM and IBD model are close to 0 (Table 3). Based on these model comparisons, we may conclude that (1) demographic models that include dispersal regulated by habitat suitability produce models that explain

Table 1. Tests of an association between genetic distances with geographic distance and/or environmental differences (as captured by two sets of environmental predictors, climate-principal component 1 (PC1) and soil-PC1) among sampling sites of individuals using distance-based redundancy analysis (see Fig. 1 for a map of sampling sites). Results are given for each geographic and environmental variable separately (i.e., the marginal tests), as well as conditioned on the effects of geographic distance (i.e., the relationship between the predictor and the response matrix controlling for geographic distance as a covariate; see text for details). Shown are the multivariate F -statistics, associated P -values, and the percentage of variance explained by each variable; significant P -values are shown in bold.

Variable	Marginal tests			Conditional tests		
	F	P -value	% Variance	F	P -value	% Variance
Distance	2.876	0.005	58.554			
Climate-PC1	8.430	0.010	9.120	1.173	0.230	2.061
Soil-PC1	1.981	0.160	2.304	0.736	0.550	1.293

the genetic divergence patterns within *L. lineopunctulata* (e.g., comparing the cENM and dENM to the unlikely IBD model); and (2) a demographic model that involves habitat shifts is much more likely to explain intraspecific divergence within *L. lineopunctulata* than static landscape models (e.g., comparing the dENM to the cENM; Table 3).

Given the dENM best explains the data, analyses were conducted to validate the accuracy of the dENM. The estimation accuracy of the four parameters differs significantly (Table 3; Fig. 4). Posterior probability of maximum carrying capacity (K_{\max}) is much flatter than the other three parameters (notice the density of the highest peak; Fig. 4) and there is limited power to estimate carrying capacity, as indicated by a $R^2 = 0.046$ (Table 3) and RMSE plot (Fig. S1). Testing of estimation bias of the parameters shows that posterior distribution of K is too narrow and that of μ is too wide (Fig. 5; histograms of the posterior quantiles significantly deviate from a uniform distribution after Bonferroni correction for multiple testing, P -value < 0.01). The other two parameters are more or less uniformly distributed so that migration rates (m) and ancestral population size (N_{Anc}) before expansion are better estimated parameters from the set of four parameters. The ancestral population of the species is estimated to be about

26,000 (Fig. 4), and the mode of the migration rate is about 0.0027 per 10 years per deme ($\sim 121 \text{ km}^2$), that is, about 3% of the population per deme emigrates in 10 years.

Discussion

Detecting spatial structure and identifying correlates of patterns of spatial structure are themselves arguably important endeavors and have received enormous attention in landscape genetics (Storfer et al. 2007, 2010; Guillot et al. 2009). What has yet to be fully explored, and remains underdeveloped, are statistical frameworks for exploring the links between such patterns and the processes that capture biological phenomena critical to addressing issues such as *how* the environment shapes patterns of genetic variation within species (Cushman and Landguth 2010; Balkenhol and Landguth 2011; Shirk et al. 2012) through the modeling of expected patterns of genetic variation.

Our study highlights the need for expanding the traditional perspective and foci of landscape genetics (as discussed later), while also presenting one approach for establishing and testing the links between pattern and process. As such, the study represents a promising new direction for expanding landscape genetic

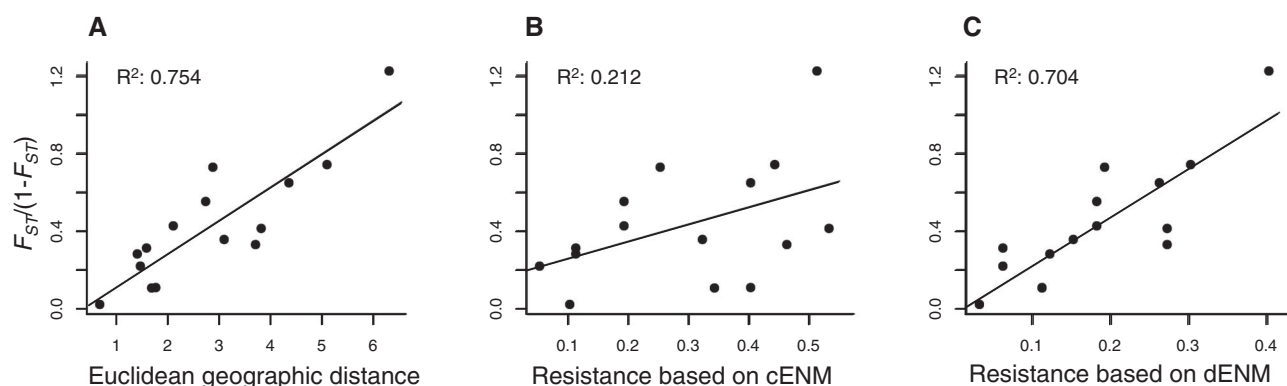


Figure 3. Plots of linearized F_{ST} against (A) pairwise population Euclidean geographic distance, and pairwise population resistance calculated from (B) contemporary habitat suitability; and (C) the composite suitability of past and present habitats (see text for details). Fitted line of the points and its R^2 are also shown.

Table 2. Results of isolation-by-resistance as calculated using Mantel and partial Mantel tests (with geography and the current ecological niche model [ENM] as covariates) between the pairwise F_{ST} -values with geographic distances and resistance matrices (i.e., rescaled geographic distances according to the suitability of habitats) separating populations (see also Fig. 3). Two resistance matrices are tested: the first is calculated from current habitat suitability score, and the second from the average of past and current suitability. Correlation coefficients (r) and the P -values from 1000 permutation tests are shown. In partial Mantel tests, covariates are listed on the second row; significant tests are shown in bold.

Matrices	Mantel tests		Partial Mantel tests			
	r	P-value	Geography		Current ENM	
			r	P -value	r	P -value
Average pairwise Euclidean distance	0.868	0.005	–	–	–	–
Resistance-values calculated from a map of habitat suitabilities from the current ENM	0.460	0.050	– 0.729	0.007	–	–
Resistance-values calculated from a composite map of habitat suitabilities from current and past ENMs	0.839	0.010	0.024	0.499	0.892	0.008

study. Nevertheless, we also recognize aspects of such complex models that would greatly benefit from further attention. These are discussed with the intention of motivating future development, but also drawing attention to aspects of the analyses that should be interpreted cautiously.

IMPORTANCE OF EXPLORING THE LINKS BETWEEN GENETIC PATTERNS AND PROCESS IN LANDSCAPE GENETICS

Different demographic processes may lead to the same genetic patterns (Csillery et al. 2010). As a consequence, an explicit model

Table 3. Properties of models and the prior and posterior distributions of estimated parameters. Bayes factor is the ratio between the highest marginal density among models and that of each model. K_{\max} = carrying capacity of the deme with highest suitability; m = migration rate per deme per generation; μ = average mutation rate; N_{Anc} = ancestral population size before expansion from the refugia. Logarithmic of all priors are uniformly distributed and have the same prior ranges across models. R^2 = coefficient of determination between a parameter and the six used partial least square components, show the power of estimating certain parameters. HPDI 50 and 90 are the interval of 50% and 90% parameter regions with the highest posterior density.

Models	Marginal density		Parameters	Prior [min, max]	R^2	Posterior		
	(P -value)	Bayes factor				Mode	HPDI 50	HPDI 90
IBD	2.14×10^{-14} (<0.0002)	9.12×10^8	$\log_{10}(K_{\max})$	[3, 5.3]	0.080	3.465	[3.256, 3.697]	[3.000, 4.301]
			$\log_{10}(m)$	[–4, –0.3]	0.012	– 3.290	[–3.701, –2.767]	[–4.000, –1.832]
			$\log_{10}(\mu)$	[–8, –6]	0.223	– 6.040	[–6.061, –6.000]	[–6.162, –6.000]
			$\log_{10}(N_{\text{Anc}})$	[3, 5]	0.546	3.848	[3.768, 3.929]	[3.626, 4.071]
cENM	5.82×10^{-8} (0.0216)	334.72	$\log_{10}(K_{\max})$	[3, 5.3]	0.145	3.000	[3.000, 3.116]	[3.000, 3.279]
			$\log_{10}(m)$	[–4, –0.3]	0.045	– 3.851	[–4.000, –3.589]	[–4.000, –3.028]
			$\log_{10}(\mu)$	[–8, –6]	0.910	– 6.000	[–6.061, –6.000]	[–6.061, –6.000]
			$\log_{10}(N_{\text{Anc}})$	[3, 5]	0.744	3.929	[3.849, 4.030]	[3.667, 4.192]
dENM	1.95×10^{-5} (0.1514)	–	$\log_{10}(K_{\max})$	[3, 5.3]	0.046	4.975	[4.487, 5.230]	[3.604, 5.300]
			$\log_{10}(m)$	[–4, –0.3]	0.541	– 1.571	[–1.870, –1.309]	[–2.243, –0.898]
			$\log_{10}(\mu)$	[–8, –6]	0.915	– 6.242	[–6.343, –6.142]	[–6.505, –6.020]
			$\log_{10}(N_{\text{Anc}})$	[3, 5]	0.737	4.414	[4.293, 4.535]	[4.131, 4.717]

cENM = contemporary ecological niche model; dENM = dynamic ecological niche model; IBD = isolation-by-distance.

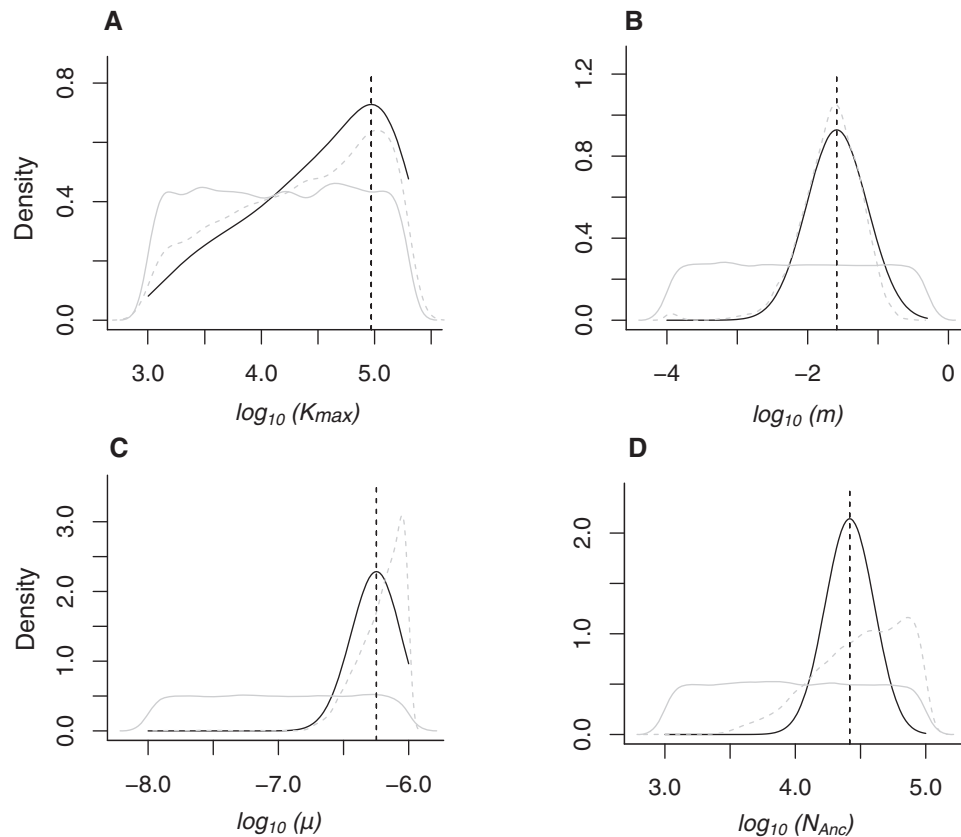


Figure 4. Posterior distribution (shown as dark line) and mode (i.e., the vertical line) of parameter estimates for the most probable model—the dENM—based on a general linear model regression adjustment of the 5000 closest simulations (see text for detail). The distribution of retained simulations (shown as dashed line) and the prior (shown as gray line) are given to highlight: (i) the improvement the general linear model procedure had on the parameter estimates (i.e., contrasting the dashed and solid dark lines); and (ii) that the data contained information relevant to estimating the parameters (i.e., contrast the solid dark and gray lines).

that can generate explicit patterns of genetic variation under different scenarios is critical (Knowles 2008, 2009). For example, patterns of shared polymorphism may not necessarily reflect recent hybridization (Green et al. 2010), but instead be a case of incomplete lineage sorting, especially if ancient population structure contributes to longer coalescent times (Eriksson and Manica

2012). Similarly, although geographic barriers might generate substantial genetic differentiation among populations (Knowles 2001), such a pattern might be generated without geographic isolation through the colonization process associated with climate-induced habitat shifts (e.g., Knowles and Alvarado-Serrano 2010). With explicit modeling, a spectrum of factors capable of

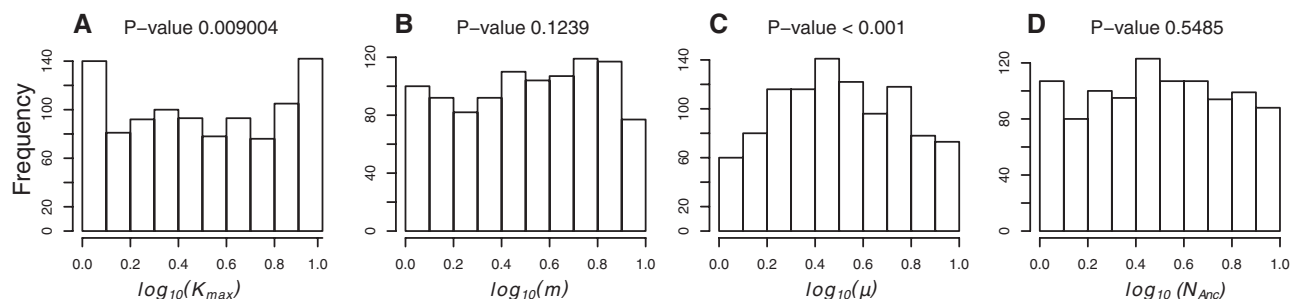


Figure 5. Distribution of posterior quantiles of parameters for the most probable model—the dENM—for evaluating potential bias in the parameter estimates, as measured by a departure from a uniform distribution using a Kolmogorov-Smirnov test; analyses are based on 1000 pseudo-observations. Estimation of m and N_{Anc} seem to be unbiased whereas posterior distribution of K is too narrow and that of μ is too wide.

producing the observed genetic patterns can be explored, thereby cautioning against interpretations without considering alternative processes.

As this study demonstrates, the essential importance of modeling the link between genetic pattern and process also extends to interpretations of the role of the environment in shaping patterns of genetic variation. This is perhaps exemplified best by contrasting the conclusions that might have been drawn from the descriptive association between genetic and geographic distance with the actual likelihood of an IBD as the most probable model explaining the data. Specifically, a highly significant association is detected with both individual- and population-level analyses between genetic and geographic distances (see Fig. 3; Tables 1, 2). Moreover, the effects of environmental variables, whether measured from current or past climatic variables, become inconsequential after controlling for the effect of geography, reinforcing that geography may indeed be the primary determinant of patterns of genetic variation, rather than aspects of the environment. Yet, the ABC tests clearly show that genetic drift associated with geographic distance alone (i.e., the IBD model; Fig. 2) is significantly less likely than models that consider varying aspects of population connectivity as impacted by environmental variables (i.e., the cENM and dENM are both more probable; see Bayes factors in Table 3), with the dENM that takes into account shifting habitat suitability over time as the most probable. So what explains this apparent contradiction between the conclusions that might be drawn from the descriptive patterns of genetic variation (see also Edwards et al. 2012) versus the models of the actual processes involved? Could such discrepancies reflect problems with the modeling procedure, such as biases in parameter estimates? These issues and their relevance in terms of the biological implications for *L. lineopunctulata* are discussed later.

DEMOGRAPHIC MODELING AS A TOOL FOR EVALUATING AND INTERPRETING GENETIC CORRELATIONS

Many statistical analysis tools have been developed to examine the correlation between genetic variation and geographic and/or environmental factors (e.g., Mantel 1967; Smouse et al. 1986; ter Braak and Verdonschot 1995; Epperson and Li 1996; Legendre and Anderson 1999; Legendre et al. 2002; Adriaensen et al. 2003; McRae 2006; Lee and Mitchell-Olds 2011; Wang et al. 2012). Although these approaches differ with respect to their statistical power to detect important factors (Legendre and Fortin 2010), none actually model the underlying processes generating the patterns (Balkenhol et al. 2009; Meirmans 2012). Although a virtue in some respects (e.g., such approaches are generally broadly applicable and are not particularly computationally demanding), there are also inherent limitations with respect to (i) evaluating *how* such factors might produce patterns of genetic variation; or (ii)

distinguishing among alternative hypotheses about the putative factors underlying patterns of genetic variation.

The merit of model-based inferences has become widely accepted in studies of genetic data (Knowles and Carstens 2007; Knowles 2009), especially with increased knowledge about the high variance of mutational and coalescent processes (Hudson 2002). As computational constraints are overcome algorithmically and with improved computing resources, the incorporation of biological realities has become feasible. For example, methods that model genetic diversity and divergence at the same time and regress against environmental factors (Foll and Gaggiotti 2006; Faubet and Gaggiotti 2008) can be used to evaluate which environmental factors (if any) might influence genetic divergence; although such models cannot control for spatial autocorrelation among factors (in contrast to the approach used here). Moreover, the flexibility and versatility of tools for evaluating and interpreting models (e.g., with ABC: Neuenschwander et al. 2008; Itan et al. 2009; Jaquière et al. 2011) can expand the repertoire of biological models that might be considered. This includes the incorporation of factors that are typically overlooked in descriptive correlative approaches (e.g., dbRDA, PCA, MDS), such as changes in population size and/or distribution, because of difficulties with their incorporation.

This latter point, we argue, may underlie some of the apparent discrepancies in the relative importance of geography compared to environmental factors in the descriptive versus model-based approach applied here. Specifically, even rescaled distances that incorporate aspects of the environment for testing for an association between environmental differences and genetic variation make a number of implicit simplifying assumptions. For example, even though a method like McRae's (2006) isolation-by-resistance considers multiple possible paths (as opposed to the least-cost path; Adriaensen et al. 2003), and gives a weighted average of the connectivity between the populations, this approach is only valid when landscape does not change over time. However, when a habitat is less stable over time, the level of population connectivity changes depending on the impact of habitat shifts on dispersal dynamics and population sizes (see Brown and Knowles 2012). These demographic consequences that are a direct extension of the underlying environment would necessarily impact patterns of genetic variation (i.e., changes in migration probabilities and local population sizes would impact the relative probabilities of gene lineage coalescence within demes and the times to coalescence; Excoffier et al. 2009). Consequently, when we actually model the demographic process of population movements across a landscape, whether it follows an IBD model where the environment does not impact population demographic patterns or one of the alternative models in which the environment does influence migration rates and deme sizes (e.g., the cENM and dENM; Fig. 2), it is perhaps not surprising that the results from the ABC analyses

(Table 3) and the descriptive correlative analyses (Tables 1 and 2) do not match up. However, can the results from the model-based tests be trusted justifying the trade-off between the simplicity of correlative analyses for what are admittedly complex models?

MODEL INTERPRETATION, VALIDATION, AND IMPLICATIONS FOR THE FACTORS STRUCTURING GENETIC VARIATION

Model validation is very important in ABC given that ABC approximates the likelihood of models with summary statistics (Pritchard et al. 1999; Beaumont et al. 2002), unlike full likelihood-based models that uses all of the data (Kuhner et al. 1998; Beerli and Felsenstein 2001; Hey and Nielsen 2004, 2007; Kuhner 2006; Nielsen and Beaumont 2009; Hey 2010). Post-sampling adjustment, such as regression (Beaumont et al. 2002) or GLM (Leuenberger and Wegmann 2010), can pose problems when the relationship between parameters and summary statistics is extrapolated beyond the region of the observed data set. Moreover, ABC can always produce posterior distribution even if the model is wrong (Bertorelle et al. 2010).

Given the model complexity, one of the concerns was whether the data would be sufficient to discriminate among probable and relatively improbable models, as well as give unbiased parameter estimates. Nevertheless, the several approaches used to validate the models in this study suggest that the results are generally robust.

Our primary objective is to test alternative demographic models (as opposed to a focus on specific parameter values), therefore, we used standard rejection sampling scheme (Beaumont et al. 2002). Although it takes longer computational time than other methods, such as ABC-MCMC, population Monte Carlo (PMC), and adaptive PMC (Beaumont et al. 2009; Wegmann et al. 2009; Moral et al. 2012), it does not create bias among models since performance of Monte Carlo methods are sensitive to the choice of tolerance level and proposal range (Wegmann et al. 2009). To show the support of the models, comparison of marginal densities of each model, as measured by Bayes factor alone is not enough. Rather, the *P*-value of observed data under the GLM model also needs to be checked to examine the percentage of the simulated data that match the empirical data. IBD and cENM models can be easily rejected based on the Bayes Factor (Table 3). In addition, the dENM has a higher probability of generating simulations with smaller or equal likelihood than the empirical observation, compared to the cENM (see *P*-value in Table 3). In other words, even though some idiosyncratic combination of parameters can produce data sets that match the data under the cENM, the dENM has much wider parameter regions that generate data close to the observation, which is the prerequisite for accurate parameter estimations. Posterior distributions of the parameters in the two models only differ significantly in the

estimation of K_{\max} . However, K_{\max} has the least power to be informed by the PLSs in the ABC analyses (see $R^2 < 0.1$ in Table 3). The estimation of maximum carrying capacity K_{\max} and average mutation rate μ show some level of bias in estimation based on the tests of uniformity of posterior quantile distributions from pseudo-observations (Fig. 5) in that posterior distribution of K is too narrow and that of μ is too wide. Because both of the two parameters are hyper priors that control the change of a series of local parameters, it might be harder for accurate estimation (Wegmann et al. 2010). This contrasts with the two parameters, migration and ancestral population size, that are estimated well with low-average RMSE of mode (0.19 and 0.15, respectively) and not biased (Fig. 5).

We acknowledge that the models informed by the ENMs may not capture all the potential historical scenarios that might be tested. However, this is a huge improvement over simple generic models that limit biological insights (Knowles 2009; Bertorelle et al. 2010). Moreover, the class of models generated from the iDDC approach, especially the incorporation of information from past distributions, permits tests of hypotheses that could not otherwise be identified (e.g., the impact of climate-induced distributional shifts on patterns of genetic variation; see also Hugall et al. 2002; Strasburg et al. 2007; Moussalli et al. 2009; Knowles and Alvarado-Serrano 2010). There may also certainly be other aspects of ENMs that introduce error into projected species distributions (see Stockwell and Peterson 2002; Graham et al. 2004; Araújo and Guisan 2006; Phillips et al. 2006; Lozier et al. 2009). This is an active area of research and the field of ENMs will no doubt see significant advances in the near future. Again, despite these sources of errors, we argue the potential gains outweigh the negatives (which again we note, should become minimized with the advances in ENMs). With respect to *L. lineopunctulata* specifically, this includes avoiding the misleading conclusions that would have resulted from extrapolating causation from descriptive correlates (i.e., only geography was consistently identified as a primary factor structuring variation; Tables 1 and 2) or considering a limited sphere of models (i.e., IBD was the least probable model, which was only apparent with the inclusion of the additional ENM-based models; Table 3).

Finally, despite the aforementioned caveats regarding the models and estimation of parameters, the iDDC modeling procedure that infuses the coupled ENM and coalescent models (Knowles et al. 2007; Richards et al. 2007) with ABC represents an intriguing new advance beyond past applications (Heckel et al. 2005; Cushman and Landguth 2010; Knowles and Alvarado-Serrano 2010; Balkenhol and Landguth 2011; Morgan et al. 2011; Brown and Knowles 2012; Shirk et al. 2012). Moreover, what this study highlights is the synergy between more traditional landscape genetic approaches and these model-based inferences for addressing the critical issue in model-based inference—how to

identify models to be tested (Knowles 2009). Our study shows the intriguing possibilities of using the descriptive approaches from landscape genetics, which detect associations between genetic and environmental factors, for developing suites of alternative hypotheses that can be translated into models for testing with ABC.

BIOLOGICAL IMPLICATIONS AND THE IMPORTANCE OF INTEGRATING HISTORICAL AND CONTEMPORARY ENVIRONMENTS

With limited information from the lack of ecological study of *L. lineopunctulata*, this study can provide important biological insights. It is noteworthy that because of the ABC framework, we can evaluate the probability of fairly contrasting views on the population demography of this lizard species. Specifically, with endemism along the coast (Fig. 1), the relatively high F_{ST} -values (i.e., values above 0.095, Table S8, except for the comparison between the two historically stable regions, P and SB; see also Carnaval et al. 2009) could be explained by different combinations of parameter. The high divergence level could reflect the lack of migration with small population sizes (the expected pattern under an IBD model), restricted migration due to barriers associated with the contemporary habitat configuration, or colonization associated with a shifting species distribution, as a habitat specialist would track climate-induced habitat shifts. All are plausible hypotheses for *L. lineopunctulata*, an abundant subterranean lizard restricted to sandplain and dune habitats of coastal southwestern Australia, a region subject to pronounced climate shifts during the Pleistocene (Fig. 1). The genetic data suggests that *L. lineopunctulata* exhibits fairly strong habitat specialization such that (i) not only is the IBD model unlikely compared to those incorporating an environmental component, but (ii) that the species most likely tracked shifts in their habitat as climate changed from the last glacial maximum (i.e., the dENM model is more probable than the cENM). Because the dynamic model that accounts for shifting species distributions (dENM) is more probable than a static model of the contemporary landscape (the cENM), the population parameter estimates from the ABC analyses also suggest that *L. lineopunctulata* has higher ancestral population size (~26,000) and much higher migration rate (~0.03 per 10 years) than if only the contemporary landscape had been considered (contrast estimates for dENM and cENM in Table 3). This could have ramifications for developing effective conservation management plans, supporting initiatives for preserving the processes contributing to genetic divergence (Moritz and Faith 1998).

As a recognized biological hotspot (Cincotta et al. 2000; Myers et al. 2000), our findings provide some valuable perspective on not only the factors promoting divergence within the focal species, but perhaps also those promoting diversification. A combination of an expanded sandplain habitat caused by late-

Quaternary sea-level fluctuations, local geological activity, and climate-induced distributional shifts are postulated to have driven diversification of the southwest Australian herpetofauna (Storr and Harold 1978, 1980; Hopper and Gioia 2004; Rabosky et al. 2004; Edwards 2007; Melville et al. 2008). Yet, the lack of detailed spatially and temporally explicit hypotheses have made it difficult to generalize how the SW Australian fauna would have been impacted by past geologic and climatic factors. Within the geographic area of study are a number of other endemic lizard species, many of which are codistributed with *L. lineopunctulata*, but also show a variety of ecological preferences, despite occupying similar habitats (Cogger 2000). This raises the question of whether this lizard community has responded similarly to past climatic events, or whether species-specific responses have predominated (Edwards et al. 2012). It may be that *L. lineopunctulata* has a higher dispersal ability compared to other *Lerista* species, for example, which lack both forelimbs and hind limbs (Cogger 2000; Bush et al. 2007; Wilson and Swan 2008). The iDDC approach applied here could be expanded into a comparative analysis, where species-specific characteristics (e.g., differing degrees of habitat specialization or vagility) can be taken into account when testing sets of biologically informed models for landscape genetic study.

ACKNOWLEDGMENTS

We are grateful for the opportunity to share our work as part of the special volume on landscape genetics and two anonymous reviewers for valuable comments to improve the earlier version of the paper. We also appreciate the help from D. F. Alvarado-Serrano in GIS-related analysis. Financial support was provided by a National Science Foundation grant (DEB-07-15487) to L. L. Knowles.

LITERATURE CITED

- Adriaenssens, F., J. P. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulincx, and E. Matthysen. 2003. The application of 'least-cost' modelling as a functional landscape model. *Landsc. Urban. Plan.* 64:233–247.
- Araújo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33:1677–1688.
- Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18:489–522.
- Balkenhol, N., and E. L. Landguth. 2011. Simulation modelling in landscape genetics: on the need to go further. *Mol. Ecol.* 20:667–670.
- Balkenhol, N., L. P. Waits, and R. J. Dezzani. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32:818–830.
- Beaumont, M. A., W. Y. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
- Beaumont, M. A., J.-M. Cornuet, J. M. Marin, and C. P. Robert. 2009. Adaptive approximate Bayesian computation. *Biometrika* 86:983–990.
- Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* 98:4563–4568.

- Bertorelle, G., A. Benazzo, and S. Mona. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* 19:2609–2625.
- Bertozzi, T., K. L. Sanders, M. J. Siström, and M. G. Gardner. 2012. Anonymous nuclear loci in non-model organisms: making the most of high-throughput genome surveys. *Bioinformatics* 28:1807–1810.
- Boulesteix, A.-L., and K. Strimmer. 2007. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.* 8:32–44.
- Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 26:211–252.
- Brown, J. L., and L. L. Knowles. 2012. Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Mol. Ecol.* 21:3757–3775.
- Bruggeman, D. J., T. Wiegand, and N. Fernandez. 2010. The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Mol. Ecol.* 19:3679–3691.
- Bush, B., B. Maryan, R. Browne-Cooper, and D. Robinson. 2007. Reptile and frogs in the bush: southwestern Australia. University of Western Australia Press, Crawley, Western Australia.
- Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz. 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* 323:785–789.
- Carstens, B. C., and L. L. Knowles. 2006. Variable nuclear markers for *Melanoplus oregonensis* identified from the screening of a genomic library. *Mol. Ecol. Notes* 6:683–685.
- Cincotta, R. P., J. Wisniewski, and R. Engelman. 2000. Human population in the biodiversity hotspots. *Nature* 404:990–992.
- Cogger, H. G. 2000. Reptiles and amphibians of Australia, 5th ed. Reed New Holland, Sydney.
- Cook, S. R., A. Gelman, and D. B. Rubin. 2006. Validation of software for Bayesian models using posterior quantiles. *J. Comput. Graph. Stat.* 15:675–692.
- Csillery, K., M. G. B. Blum, O. E. Gaggiotti, and O. Francois. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25:410–418.
- Curat, M., and L. Excoffier. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol.* 2:2264–2274.
- Curat, M., N. Ray, and L. Excoffier. 2004. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes* 4:139–142.
- Cushman, S. A., and E. L. Landguth. 2010. Spurious correlations and inference in landscape genetics. *Mol. Ecol.* 19:3592–3602.
- Edwards, D. L. 2007. Biogeography and speciation of a direct developing frog from the coastal arid zone of Western Australia. *Mol. Phylogenet. Evol.* 45:494–505.
- Edwards, D. L., J. S. Keogh, and L. L. Knowles. 2012. Effects of vicariant barriers, habitat stability, population isolation and environmental features on species divergence in the south-western Australian coastal reptile community. *Mol. Ecol.* 21:3809–3822.
- Epperson, B. K., and T. Q. Li. 1996. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proc. Natl. Acad. Sci. USA* 93:10528–10532.
- Epperson, B. K., B. H. McRae, K. Scribner, S. A. Cushman, M. S. Rosenberg, M.-J. Fortin, P. M. A. James, M. Murphy, S. Manel, P. Legendre, et al. 2010. Utility of computer simulations in landscape genetics. *Mol. Ecol.* 19:3549–3564.
- Eriksson, A., and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl. Acad. Sci. USA* 109:13956–13960.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol. Ecol. Res.* 10:564–567.
- Excoffier, L., J. Novembre, and S. Schneider. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J. Hered.* 91:506–509.
- Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* 40:481–501.
- Faubet, P., and O. E. Gaggiotti. 2008. A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* 178:1491–1504.
- Foll, M., and O. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. *Genetics* 174:875–891.
- Gompert, Z., M. L. Forister, J. A. Fordyce, C. C. Nice, R. J. Williamson, and C. Alex Buerkle. 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Mol. Ecol.* 19:2455–2473.
- Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781–1793.
- Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. W. Zhai, M. H. Y. Fritz, et al. 2010. A draft sequence of the neandertal genome. *Science* 328:710–722.
- Guillot, G., R. Leblois, A. Coulon, and A. C. Frantz. 2009. Statistical methods in spatial genetics. *Mol. Ecol.* 18:4734–4756.
- Heckel, G., R. Burri, S. Fink, J. F. Desmet, and L. Excoffier. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution* 59:2231–2242.
- Hey, J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol.* 27:905–920.
- Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.
- . 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. USA* 104:2785–2790.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Intl. J. Climatol.* 25:1965–1978.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83:2027–2036.
- Hocking, R. M., H. T. Moors, and W. J. E. Van de Graaff. 1987. Geology of the Carnarvon basin, Western Australia. State Print. Division, Perth, Australia.
- Hopper, S. D., and P. Gioia. 2004. The southwest Australian floristic region: evolution and conservation of a global hot spot of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 35:623–650.
- Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Hugall, A., C. Moritz, A. Moussalli, and J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proc. Natl. Acad. Sci. USA* 99:6112–6117.
- Hull, J. M., A. C. Hull, B. N. Sacks, J. P. Smith, and H. B. Ernest. 2008. Landscape characteristics influence morphological and genetic differentiation in a widespread raptor (*Buteo jamaicensis*). *Mol. Ecol.* 17:810–824.

- Itan, Y., A. Powell, M. A. Beaumont, J. Burger, and M. G. Thomas. 2009. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* 5:e1000491.
- Jacquière, J., T. Broquet, A. H. Hirzel, J. Yearsley, and N. Perrin. 2011. Inferring landscape effects on dispersal from genetic distances: how far can we go? *Mol. Ecol.* 20:692–705.
- Jeffreys, H. 1961. *Theory of probability*, 3rd ed. Clarendon Press, Oxford, U.K.
- Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service. *BMC Genet.* 6:13.
- Knowles, L. L. 2001. Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Mol. Ecol.* 10:691–701.
- . 2008. Why does a method that fails continue to be used? *Evolution* 62:2713–2717.
- . 2009. Statistical phylogeography. *Annu. Rev. Ecol. Evol. Syst.* 40:593–612.
- Knowles, L. L., and D. F. Alvarado-Serrano. 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Mol. Ecol.* 19:3727–3745.
- Knowles, L. L., and B. C. Carstens. 2007. Estimating a geographically explicit model of population divergence. *Evolution* 61:477–493.
- Knowles, L. L., B. C. Carstens, and M. L. Keat. 2007. Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr. Biol.* 17:940–946.
- Kuhner, M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770.
- Kuhner, M. K., J. Yamato, and J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429–434.
- Landguth, E. L., S. A. Cushman, M. A. Murphy, and G. Luikart. 2010. Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Mol. Ecol. Res.* 10:854–862.
- Lee, C. R., and T. Mitchell-Olds. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol. Ecol.* 20:4631–4642.
- Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* 69:1–24.
- Legendre, P., and M.-J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Res.* 10:831–844.
- Legendre, P., M. R. T. Dale, M. J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25:601–615.
- Leuenberger, C., and D. Wegmann. 2010. Bayesian computation and model selection without likelihoods. *Genetics* 184:243–252.
- Lozier, J. D., P. Aniello, and M. J. Hickerson. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *J. Biogeogr.* 36:1623–1627.
- Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921–931.
- Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18:189–197.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.
- McKenzie, N. J., D. W. Jacquier, L. J. Ashton, and H. P. Cresswell. 2000. Estimation of soil properties using the atlas of Australian soils. Technical Report 11/00, CSIRO Land and Water, Canberra.
- McRae, B. H. 2006. Isolation by resistance. *Evolution* 60:1551–1561.
- Meirmans, P. G. 2012. The trouble with isolation by distance. *Mol. Ecol.* 21:2839–2846.
- Melville, J., L. P. Shoo, and P. Doughty. 2008. Phylogenetic relationships of the heath dragons (*Rankinia adelaidensis* and *R. parviceps*) from the south-western Australian biodiversity hotspot. *Aust. J. Zool.* 56:159–171.
- Mendez, M., H. C. Rosenbaum, A. Subramaniam, C. Yackulic, and P. Bordino. 2010. Isolation by environmental distance in mobile marine species: molecular ecology of franciscana dolphins at their southern range. *Mol. Ecol.* 19:2212–2228.
- Mevik, B.-H., and R. Wehrens. 2007. The PLS package: principal component and partial least squares regression in R. *J. Stat. Softw.* 18:1–24.
- Moral, P. D., A. Doucet, and A. Jasra. 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.* 22:1009–1020.
- Morgan, K., S. M. O’Loughlin, B. Chen, Y. M. Linton, D. Thongwat, P. Somboon, M. Y. Fong, R. Butlin, R. Verity, A. Prakash, et al. 2011. Comparative phylogeography reveals a shared impact of Pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Mol. Ecol.* 20:4533–4549.
- Moritz, C., and D. P. Faith. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. *Mol. Ecol.* 7:419–429.
- Mory, A. J., R. P. Iasky, and K. A. R. Ghorri. 2003. A summary of the geological evolution and the petroleum potential of the southern carmarvon basin western australia. Department of Industry and Resources, Government of Western Australia, Perth, Australia.
- Moussalli, A., C. Moritz, S. E. Williams, and A. C. Carnaval. 2009. Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Mol. Ecol.* 18:483–499.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.
- Neuenschwander, S., C. R. Lurgiader, N. Ray, M. Currat, P. Vonlanthen, and L. Excoffier. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol. Ecol.* 17:757–772.
- Nielsen, R., and M. A. Beaumont. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* 18:1034–1047.
- Northcote, K. H., G. G. Beckmann, E. Bettenay, H. M. Churchward, D. C. Van Dijk, G. M. Dimmock, G. D. Hubble, R. F. Isbell, W. M. McArthur, G. G. Murtha, et al. 1960–1968. Atlas of Australian soils, sheets 1 to 10. with explanatory data. CSIRO Australia and Melbourne Univ. Press, Melbourne, Australia.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, et al. 2012. Vegan: community ecology package. R package version 2.0-3. Available at <http://CRAN.R-project.org/package=vegan>. Accessed June 10, 2012.
- Perrier, C., R. Guyomard, J. L. Bagliniere, and G. Evanno. 2011. Determinants of hierarchical genetic structure in Atlantic salmon populations: environmental factors vs. anthropogenic influences. *Mol. Ecol.* 20:4231–4245.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araújo. 2011. Ecological niches and geographic distributions. Monographs in population biology 49. Princeton Univ. Press, Princeton, NJ.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190:231–259.

- Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–1798.
- R Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rabosky, D. L., K. P. Aplin, S. C. Donnellan, and S. B. Hedges. 2004. Molecular phylogeny of blindsnakes (*Ramphotyphlops*) from western Australia and resurrection of *Ramphotyphlops bicolor* (Peters, 1857). *Aust. J. Zool.* 52:531–548.
- Ray, N., M. Currat, P. Berthier, and L. Excoffier. 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.* 15:1161–1167.
- Raufaste, N., and F. Rousset. 2001. Are partial mantel tests adequate? *Evolution* 55:1703–1705.
- Richards, C. L., B. C. Carstens, and L. L. Knowles. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *J. Biogeogr.* 34:1833–1845.
- Robert, C. P., J. M. Cornuet, J. M. Marin, and N. S. Pillai. 2011. Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl. Acad. Sci. USA* 108:15112–15117.
- Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644.
- Shah, V. B., and B. H. McRae. 2008. Circuitscape: a tool for landscape ecology. Pp. 62–66. *Proceedings of the 7th python in science conference (SciPy 2008)*. Pasadena, CA.
- Shirk, A. J., S. A. Cushman, and E. L. Landguth. 2012. Simulating pattern-process relationships to validate landscape genetic models. *Intl. J. Ecol.* 2012: 539109.
- Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple-regression and correlation extensions of the mantel test of matrix correspondence. *Syst. Zool.* 35:627–632.
- Sork, V. L., F. W. Davis, R. Westfall, A. Flint, M. Ikegami, H. F. Wang, and D. Grivet. 2010. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Mol. Ecol.* 19:3806–3823.
- Stockwell, D. R. B., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Modell.* 148:1–13.
- Storfer, A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. P. Waits. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98:128–142.
- Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits. 2010. Landscape genetics: where are we now? *Mol. Ecol.* 19:3496–3514.
- Storr, G. M., and G. Harold. 1978. Herpetofauna of the shark bay region, western australia. *Rec. Western Aust. Museum* 6:449–467.
- . 1980. Herpetofauna of the Zuytdorp coast and hinterland, Western Australia. *Rec. Western Aust. Museum* 8:359–375.
- Strasburg, J. L., M. Kearney, C. Moritz, and A. R. Templeton. 2007. Combining phylogeography with distribution modeling: multiple Pleistocene range expansions in a parthenogenetic gecko from the Australian arid zone. *PLoS One* 2:e760.
- Tajima, F., and M. Nei. 1984. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* 1:269–285.
- Tavare, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.
- ter Braak, C. J. F., and P. F. M. Verdonschot. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* 57:255–289.
- Thomson, R. C., I. J. Wang, and J. R. Johnson. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Mol. Ecol.* 19:2184–2195.
- Wang, C., S. Zöllner, and N. A. Rosenberg. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genet.* 8:e1002886.
- Wang, I. J. 2010. Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol. Ecol.* 19:2605–2608.
- Wegmann, D., and L. Excoffier. 2010. Bayesian inference of the demographic history of chimpanzees. *Mol. Biol. Evol.* 27:1425–1435.
- Wegmann, D., M. Currat, and L. Excoffier. 2006. Molecular diversity after a range expansion in heterogeneous environments. *Genetics* 174:2009–2020.
- Wegmann, D., C. Leuenberger, and L. Excoffier. 2009. Efficient approximate Bayesian computation coupled with mark chain Monte Carlo without likelihood. *Genetics* 182:1207–1218.
- Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinform.* 11:7.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Wilson, S., and G. Swan. 2008. A complete guide to reptiles of Australia, 2nd ed. New Holland Publishers, Sydney.
- Xu, J. W., M. Perez-Losada, C. G. Jara, and K. A. Crandall. 2009. Pleistocene glaciation leaves deep signature on the freshwater crab *Aegla alacalufi* in Chilean Patagonia. *Mol. Ecol.* 18:904–918.
- Zellmer, A. J., and L. L. Knowles. 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Mol. Ecol.* 18:3593–3602.

Associate Editor: K. Petren

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Root mean square error (RMSE) of parameter estimation against number of partial least squares included under four demographic models: (a) isolation-by-distance (IBD), (b) contemporary ecological niche model (cENM), and (c) dynamic ENM (dENM).

Table S1. Geographic locations of sampled individuals and their assigned population (see Fig. 1 for distributional details).

Table S2. List of nuclear loci sequenced in this study.

Table S3. Settings for NGen sequence assembler (DNASTAR) used for the 454 data set in the discovery of polymorphic loci.

Table S4. Length of each locus and sampling per populations for each locus.

Table S5. Soil properties used in the construction of soil layers for the principal component analysis analyses (for detailed description see McKenzie et al. 2000).

Table S6. Molecular indices calculated per locus and presented for each population separately (see Fig. 1 for distributional information), as well as across all populations, including heterozygosity (H) and the standard deviation (H_{sd}), the number of segregating sites (S), the number of haplotypes (K), and nucleotide diversity (π).

Table S7. List of summary statistics used in approximate Bayesian computation analyses.

Table S8. Pairwise F_{ST} of the six populations ordered from north to south (lower triangle) and the significance (upper triangle).